

Analysis of loss of function variants

Manuel A. Rivas, University of Oxford
Geuvadis RNA Sequencing Analysis Meeting - Barcelona
09/07/2012

with Tuuli Lappalainen, Micha Sammeth, Daniel MacArthur, Monkol Lek, Thomas Wieland, Andrew Kirby, Peter AC 't Hoen, Thasso Griebel, Natalja Kurbatova, Gabrielle Berthier and Geuvadis Consortium

Annotation

PREMATURE_STOP
REMOVED_START
SPLICE_SITE
DELETION_FS
INSERTION_FS
EXON_DELETION
EXON_DUPLICATION
GENE_DELETION
GENE_DUPLICATION

Main Annotation Classes

Monkol Lek, Daniel MacArthur, Manuel Rivas, Tuuli Lappalainen

Annotation

ALLELE - Alternative allele for the annotation

ANNOTATION_CLASS - Main annotation class

DOMAIN - Coding or non-coding

GENE_ID - Ensembl Gene ID

GENE_NAME - HGNC name

FRAME_MAINTAIN - Whether or not frame is maintained

ANNOTATION_SUBCLASS_SPLICE - Splice variant subclass

Quality of the LoF call - High Confidence / Low-Confidence

NMD - If the variant is expected to cause NMD. 50bp rule predictions.

PEPTIDES_TO_STOP - How many peptides to next stop

*Loss of Function relevant
information keys*

Information Keys

Monkol Lek, Daniel MacArthur, Manuel Rivas, Tuuli Lappalainen

Annotation

Loss of Function (LoF) Filters

Nonsense Variant



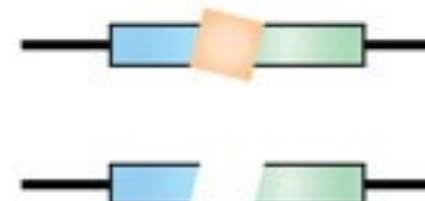
LoF allele ancestral
Flanking canonical splice sites
Exon size < 15 bp
Single exon transcript
Protein truncation < 5%

Essential Splice Site



LoF allele ancestral
Canonical splice site
Intron size < 15 bp

Frameshift Insertion/Deletions

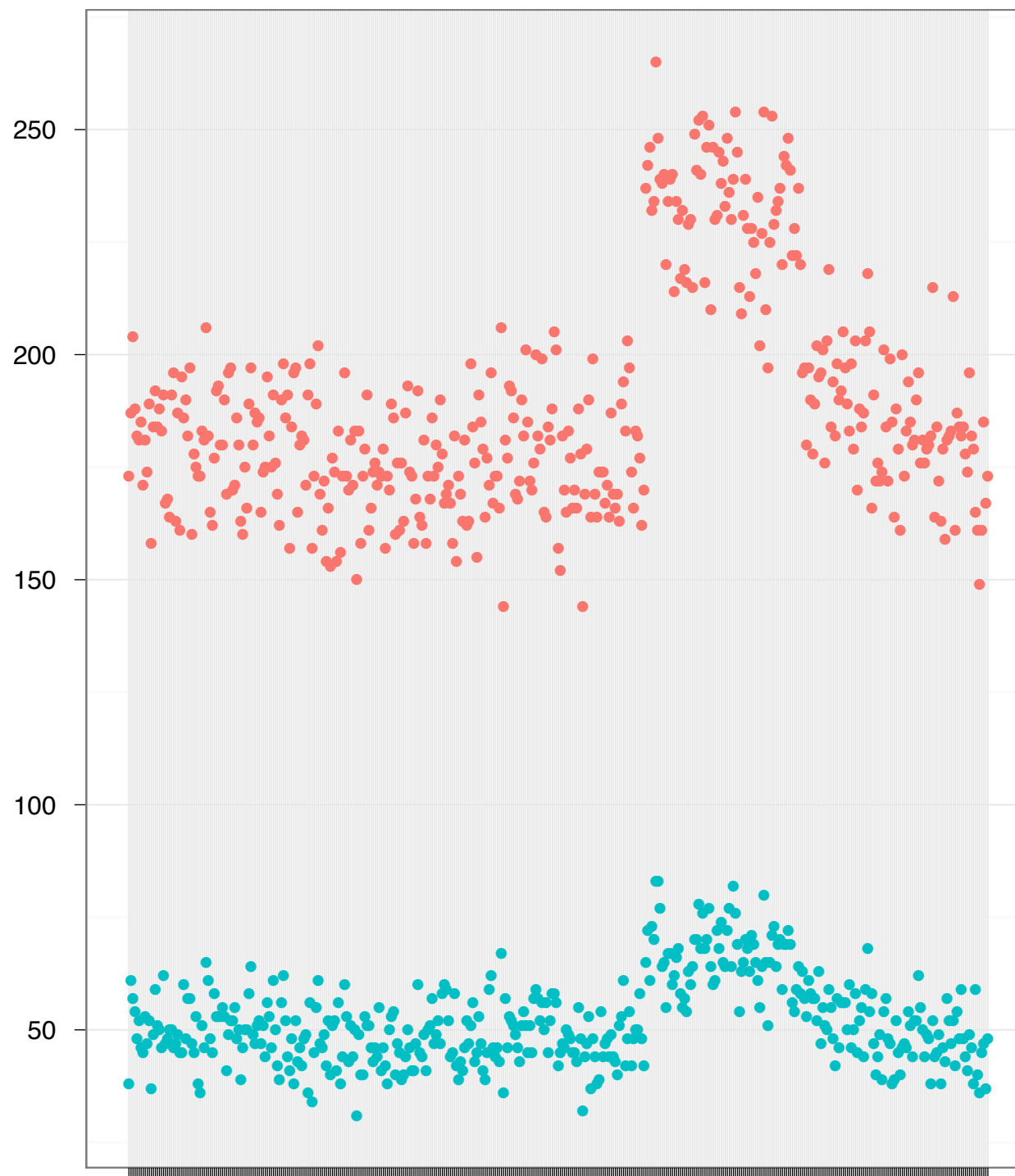


LoF allele ancestral
Flanking canonical splice sites
Exon size < 15 bp
Single exon transcript
Protein truncation < 5%

Filtering Strategies

Daniel MacArthur and Monkol Lek

Number of Heterozygote Sites



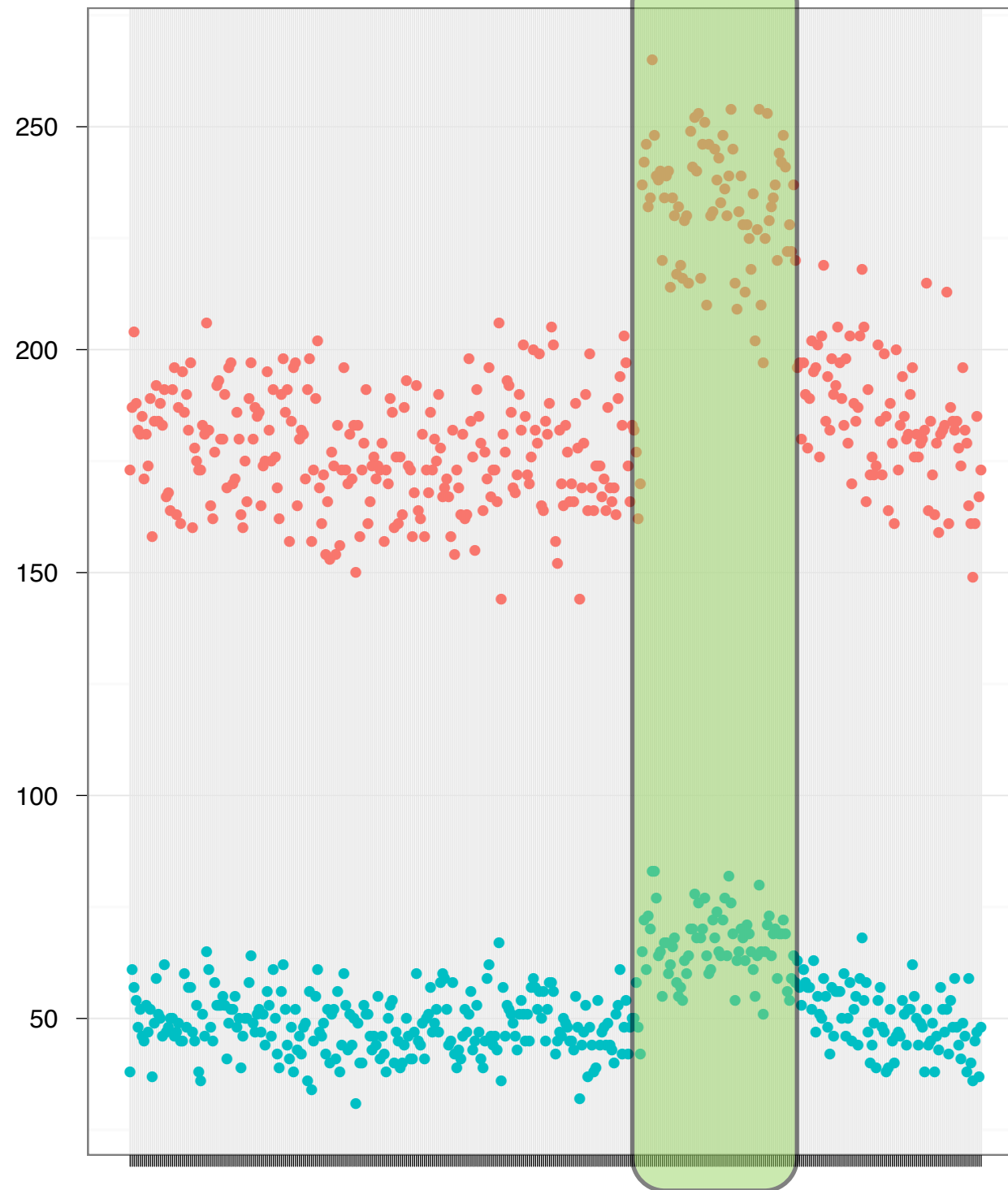
Loss of Function

STOP Gained

Subjects

Suitable for ASE Analysis

Number of Heterozygote Sites



YRI Samples

Loss of Function

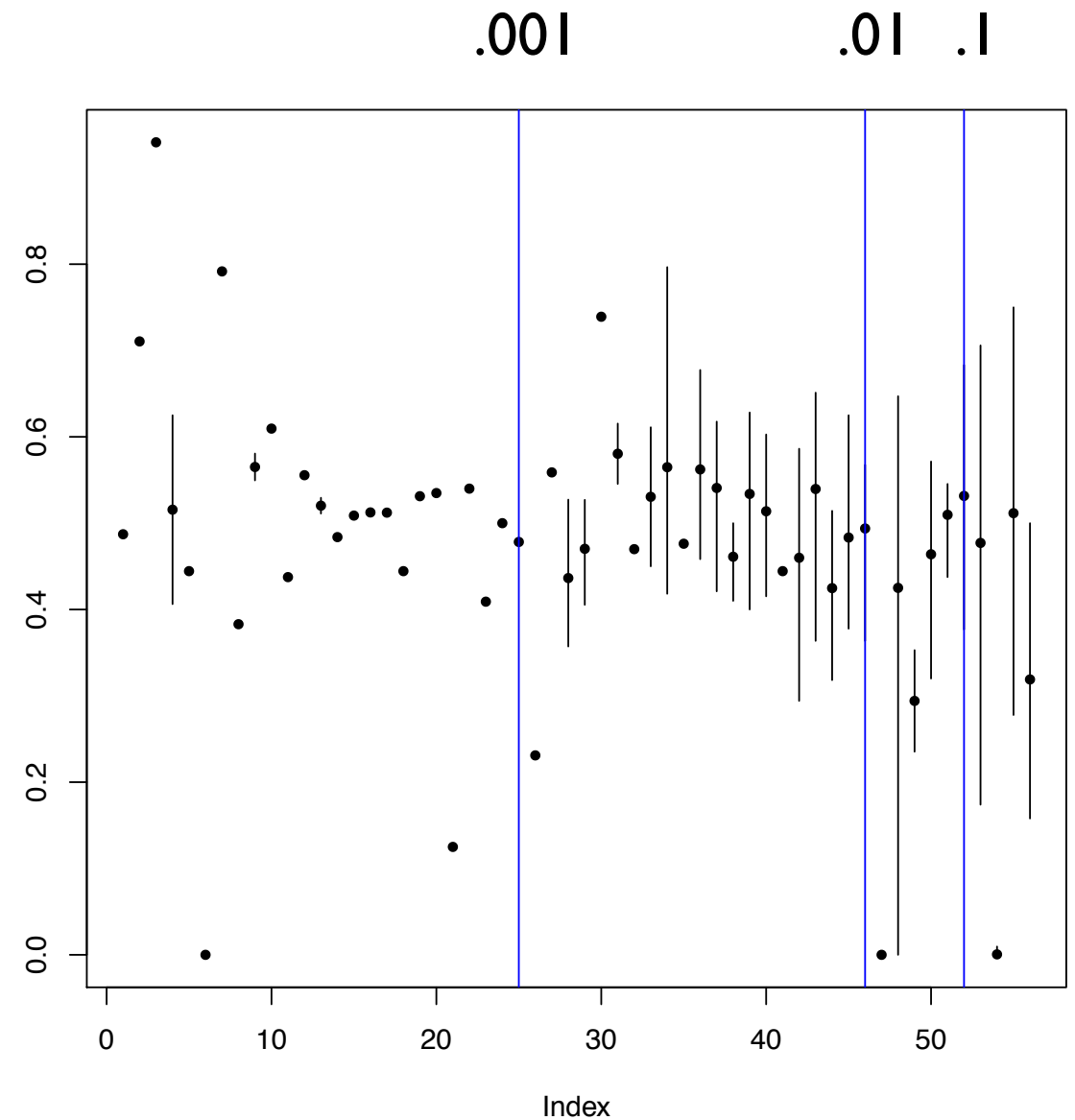
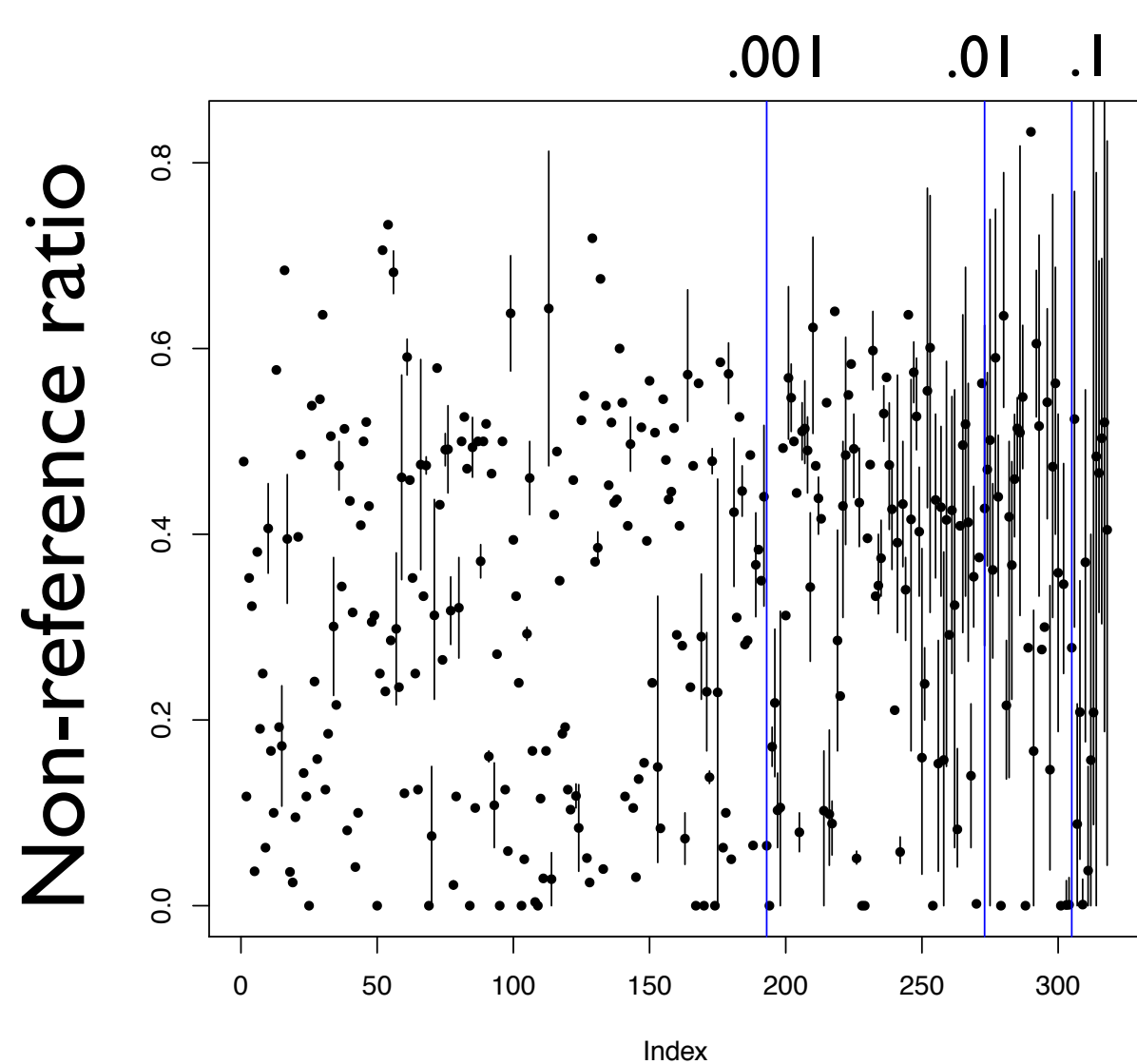
STOP Gained

Subjects

Identify QC parameters that influence the number of LoF heterozygote sites that can be analyzed

Suitable for ASE Analysis

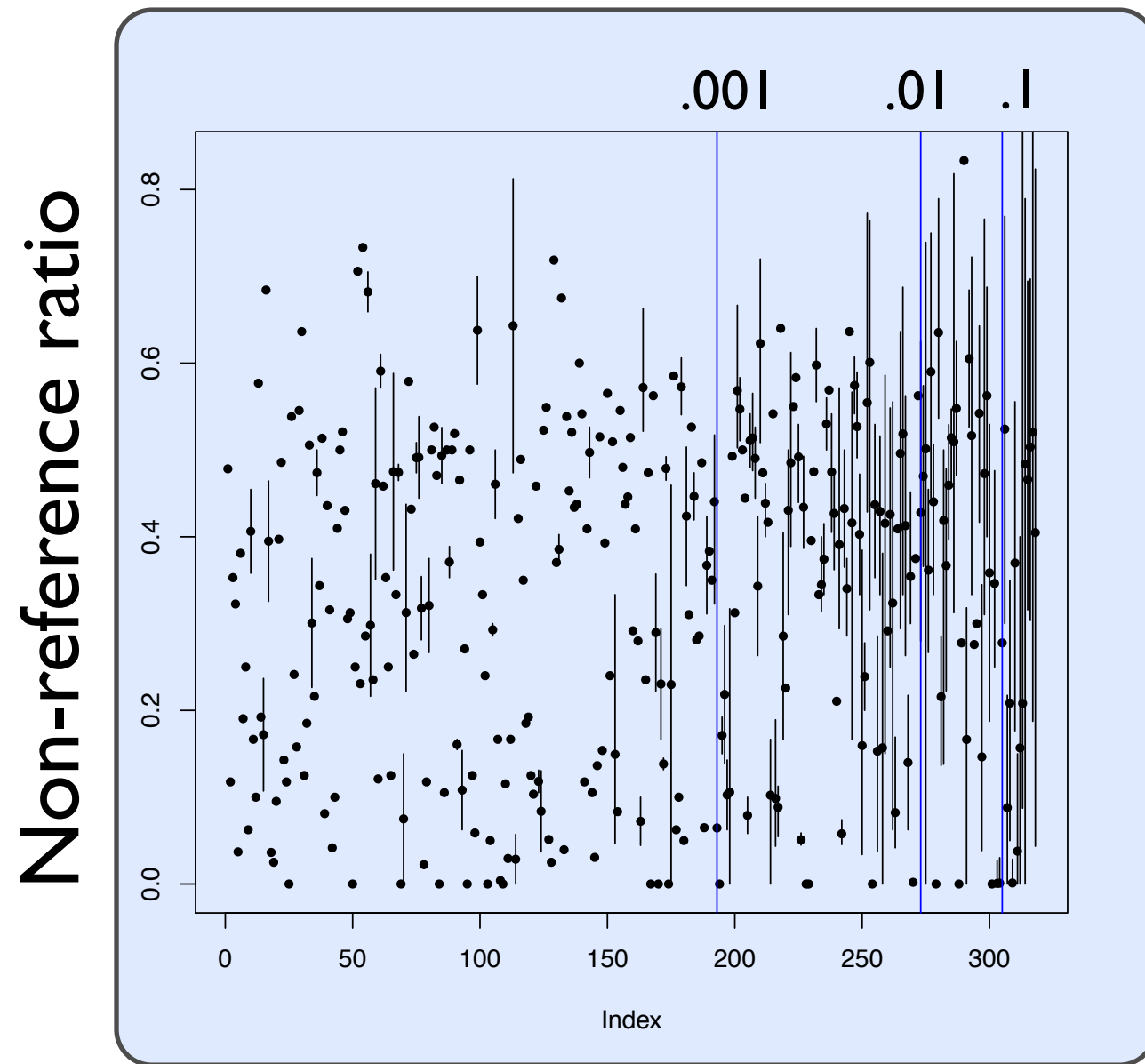
Allele Specific Expression analysis



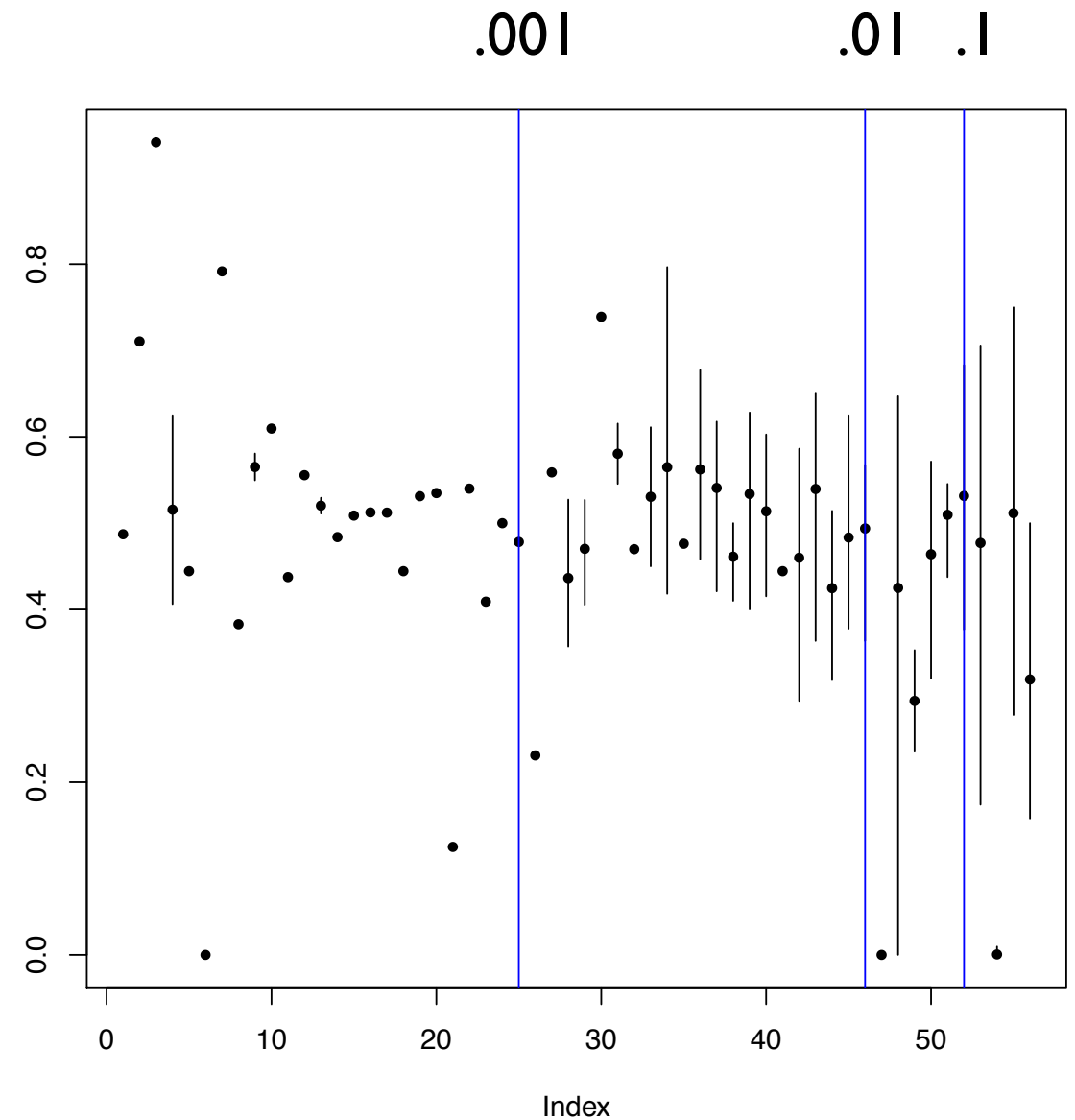
ASE Analysis

Focus on STOP GAIN variants

Allele Specific Expression analysis



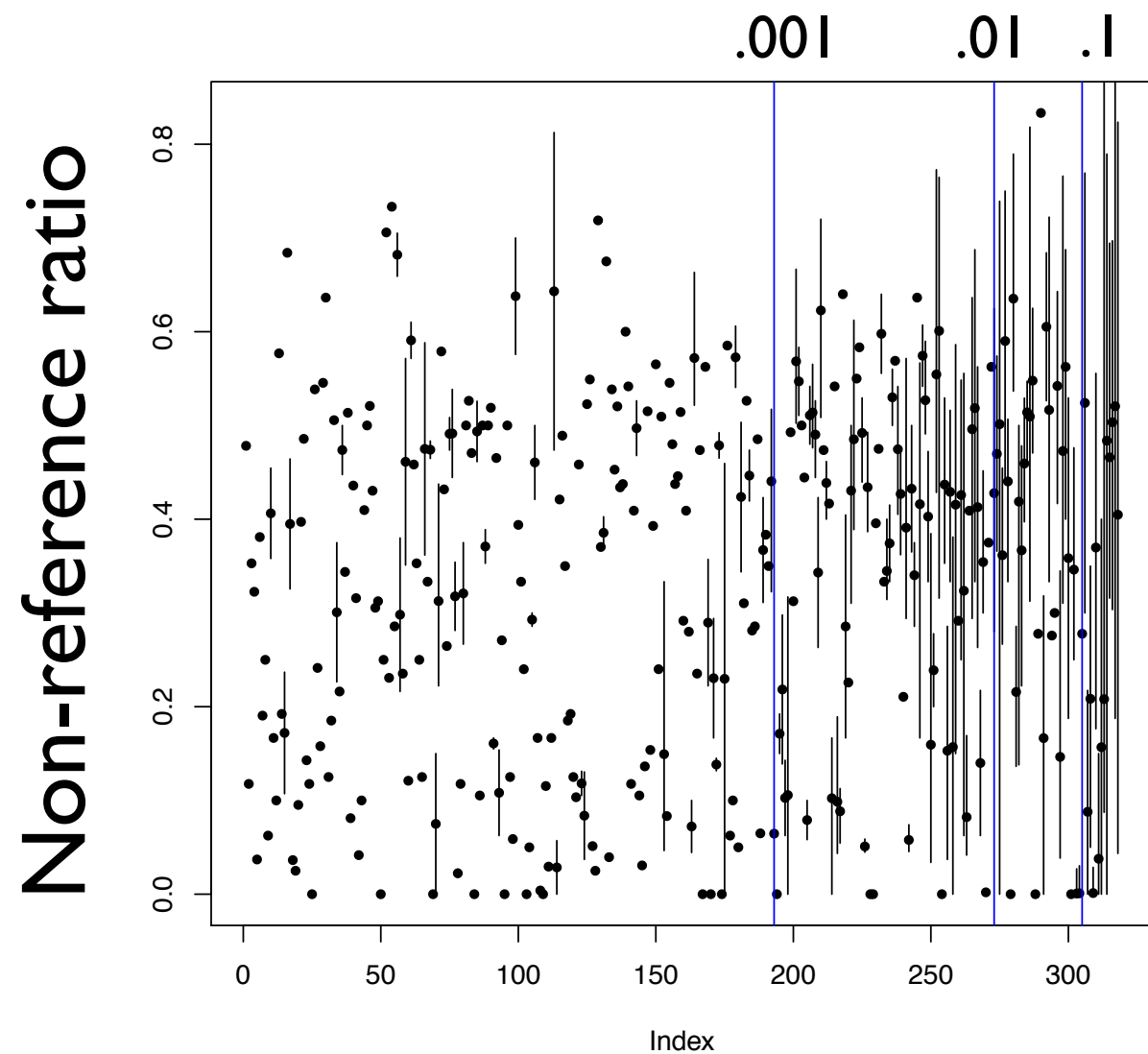
Variants Predicted to Trigger NMD



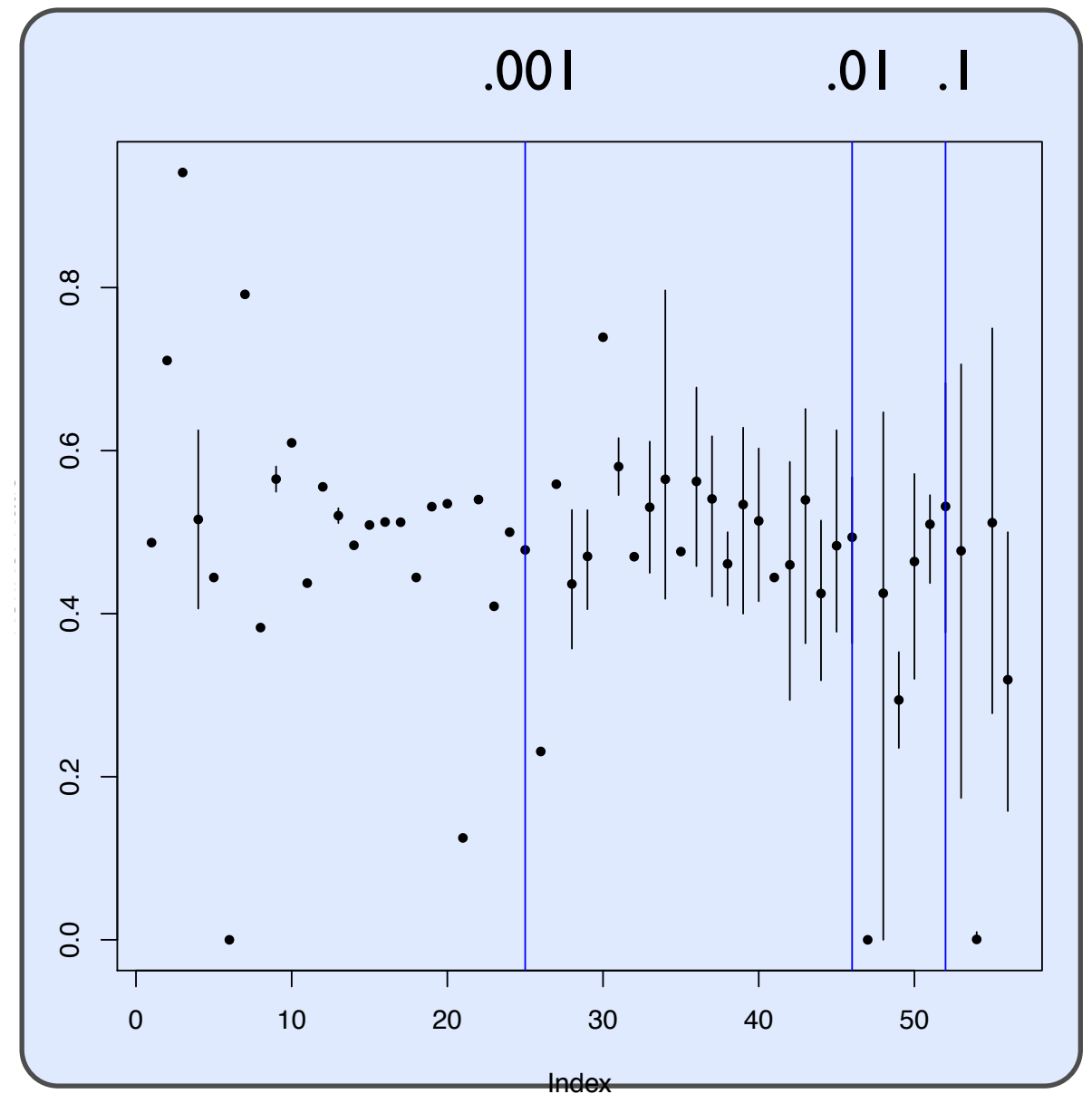
ASE Analysis

Focus on STOP GAIN variants

Allele Specific Expression analysis



Variants Predicted to Trigger NMD

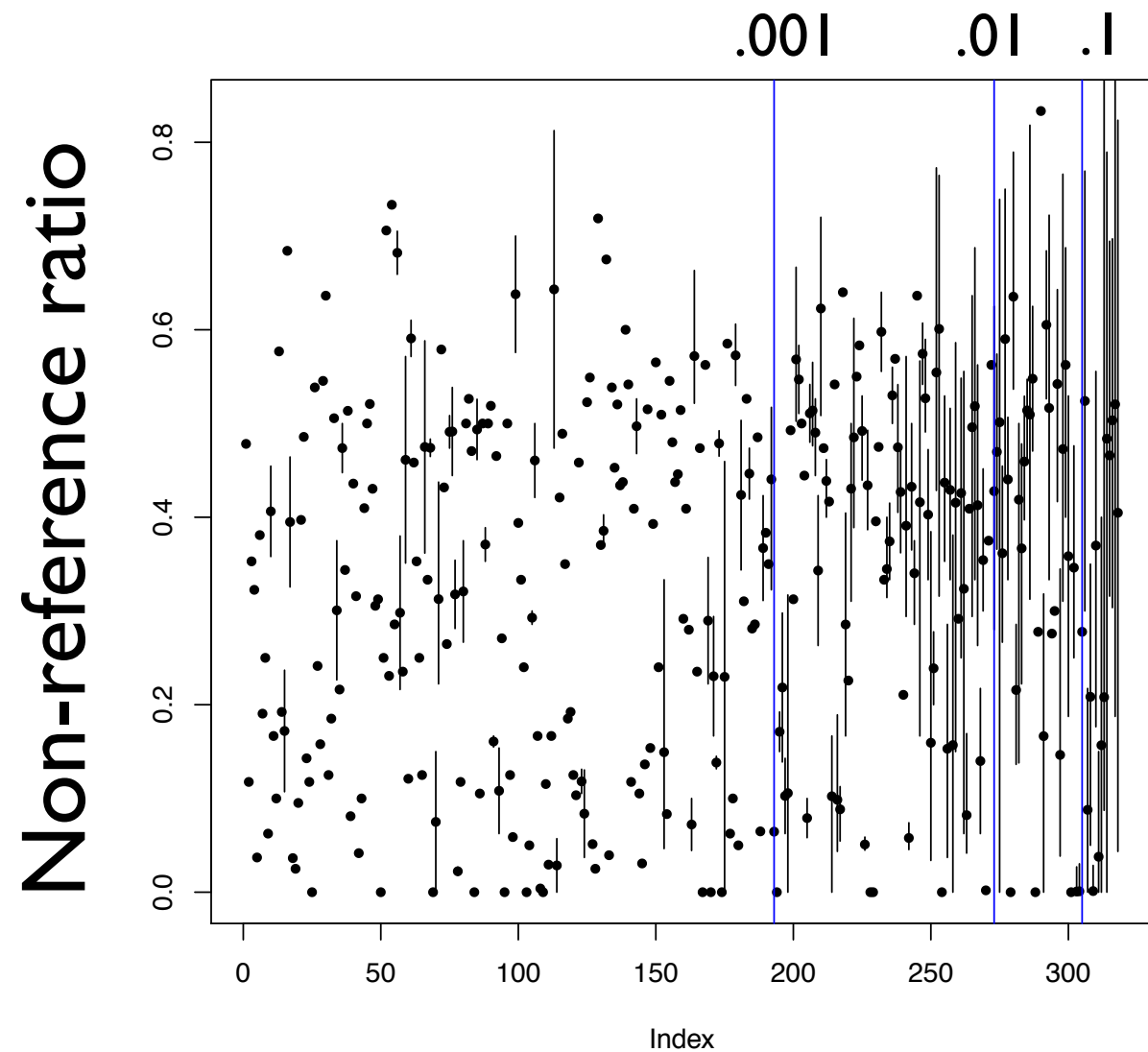


Variants Predicted to Escape NMD

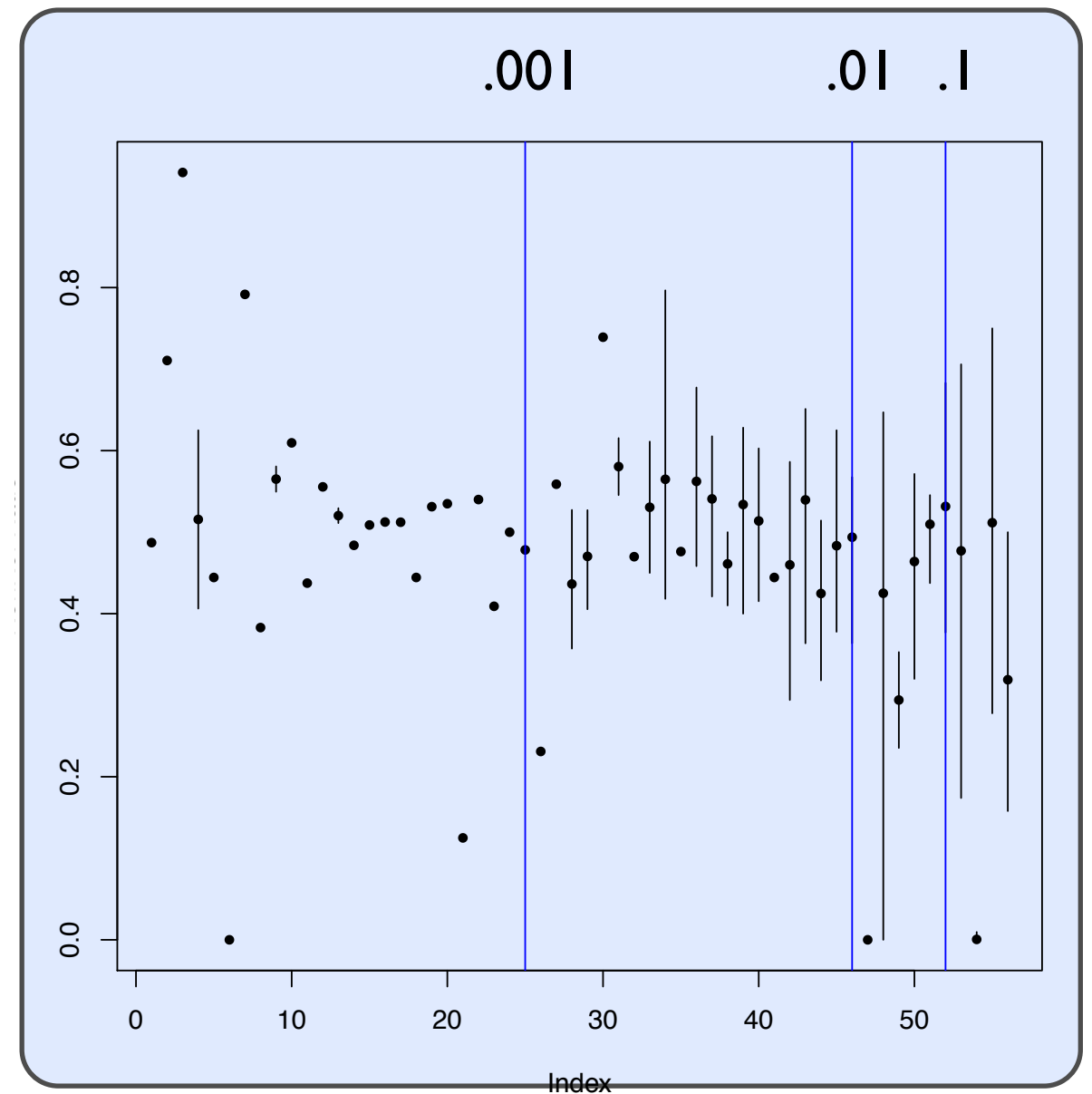
ASE Analysis

Focus on STOP GAIN variants

Allele Specific Expression analysis



Variants Predicted to Trigger NMD

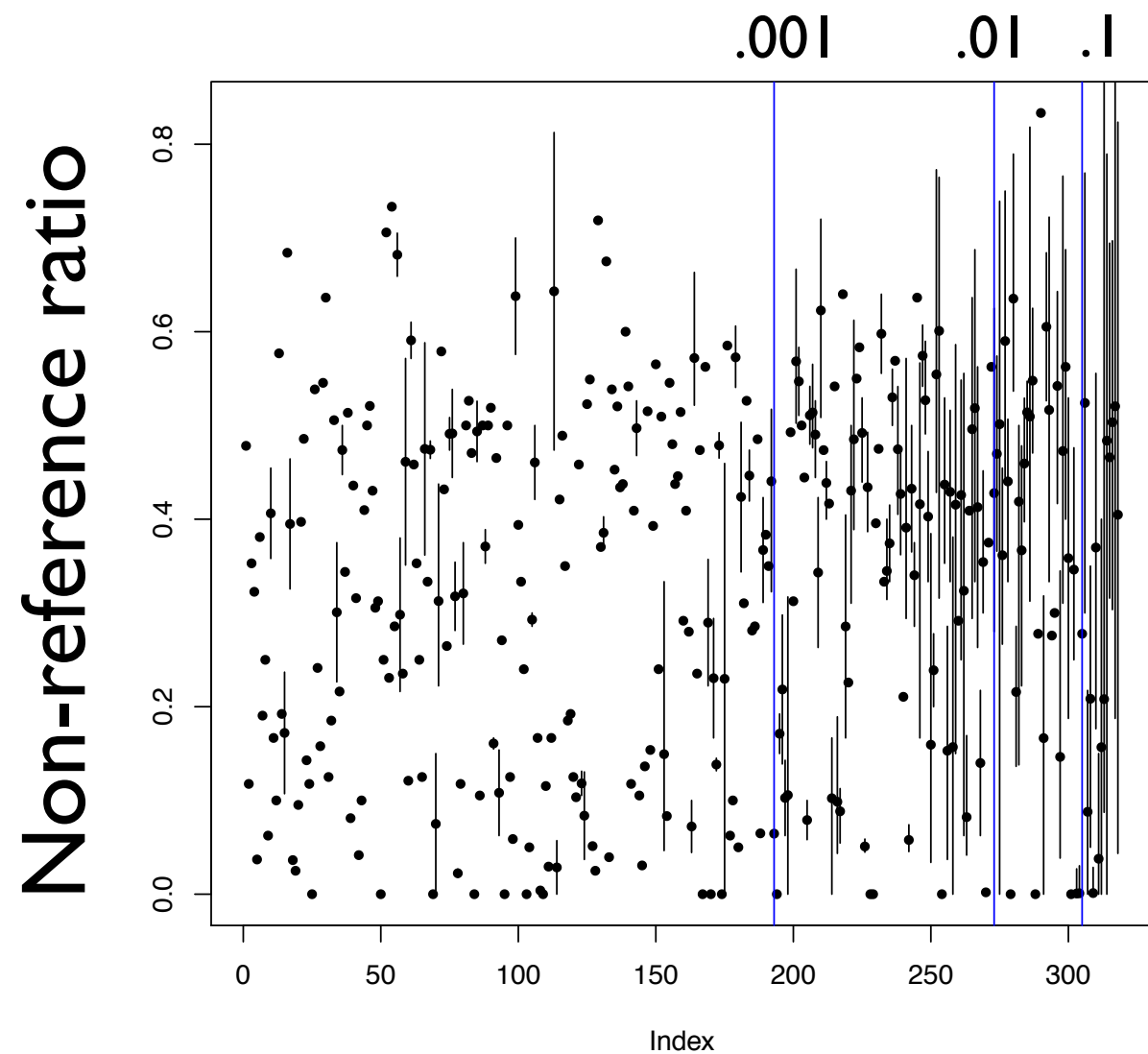


Variants Predicted to Escape NMD

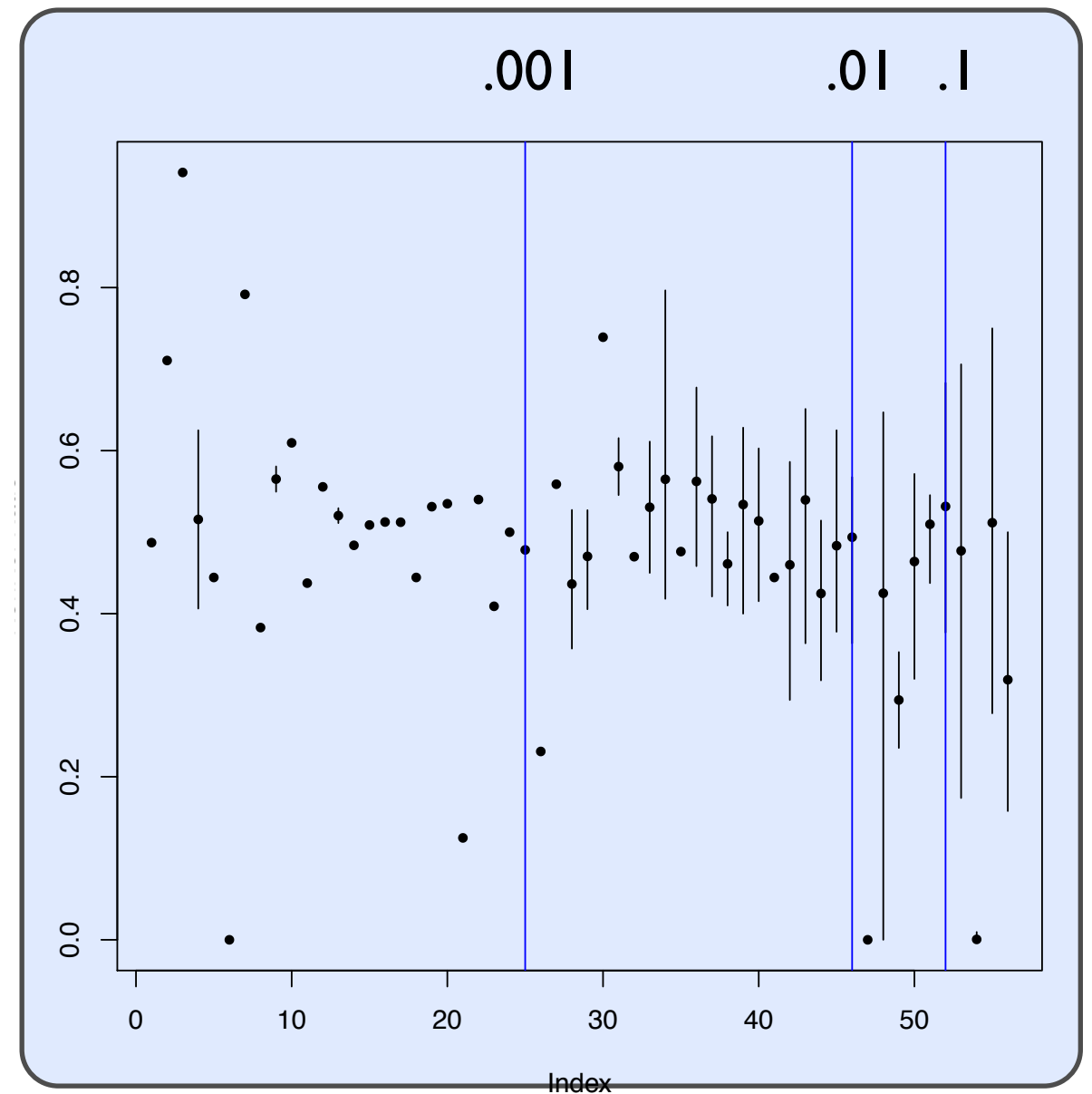
ASE Analysis

Further investigation with additional lof categories and subsequent NMD

Allele Specific Expression analysis



Variants Predicted to Trigger NMD

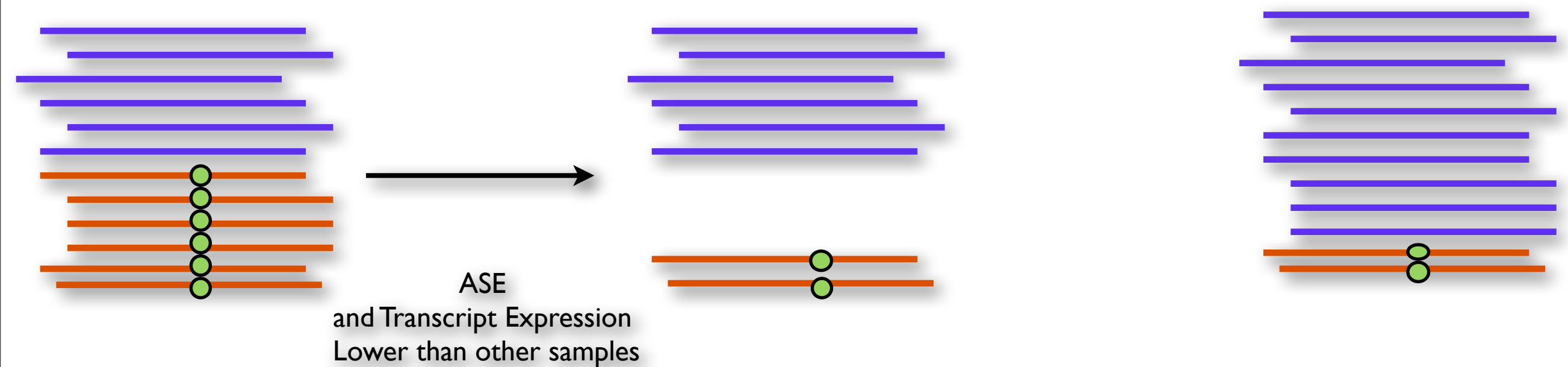


Variants Predicted to Escape NMD

ASE Analysis

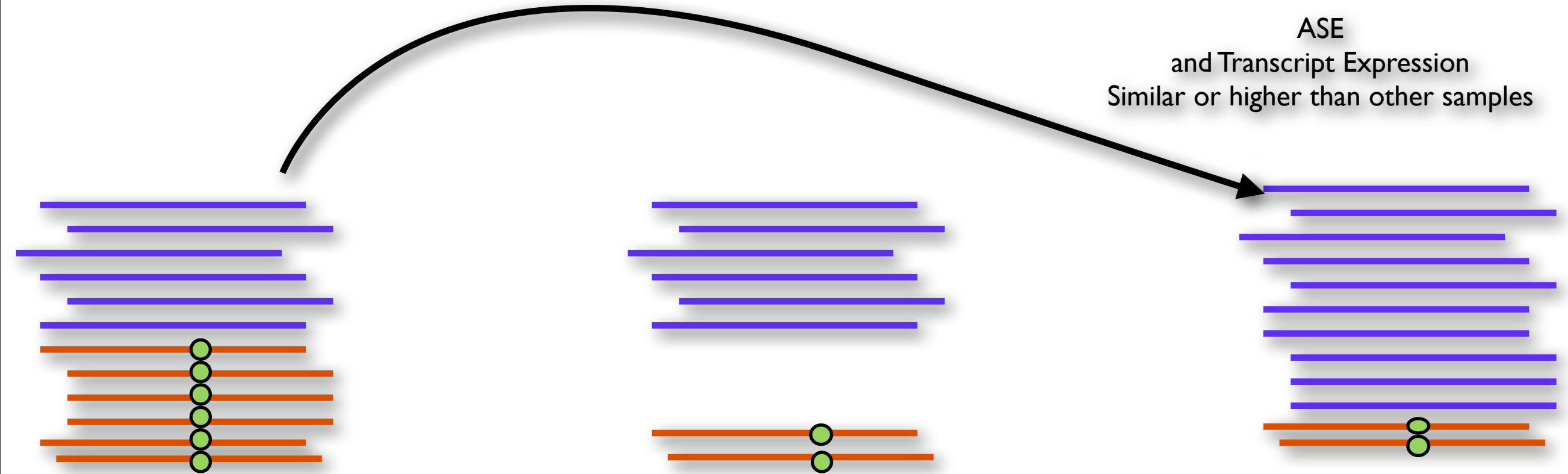
Further investigation into NMD efficiency - is there inter individual variation?

Dosage Compensation



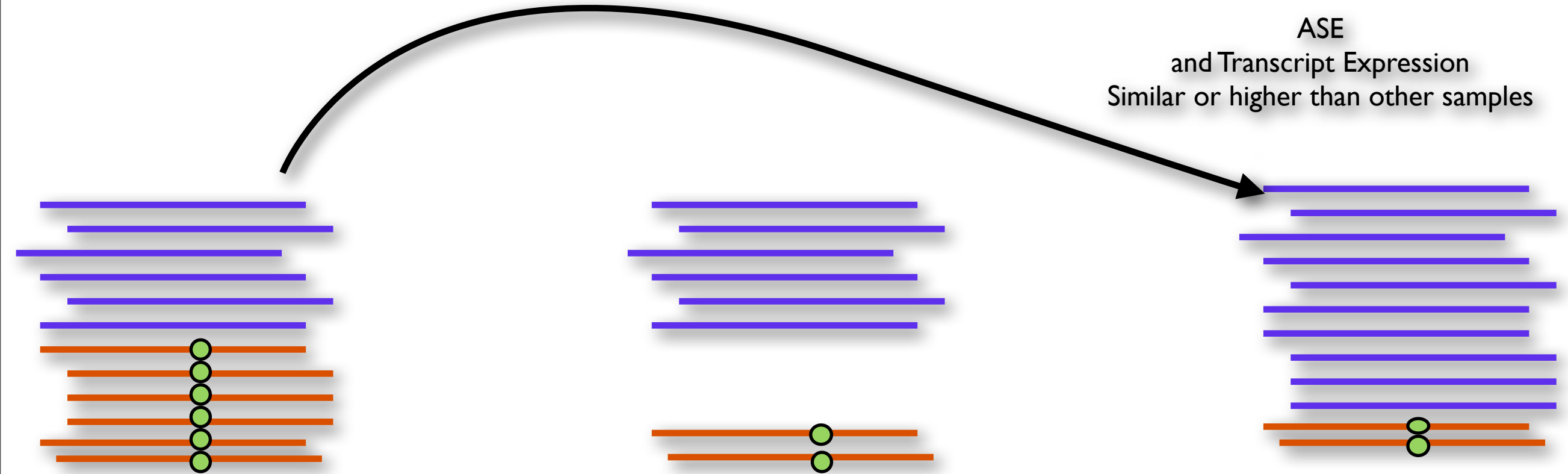
Dosage Compensation

Dosage Compensation



Dosage Compensation

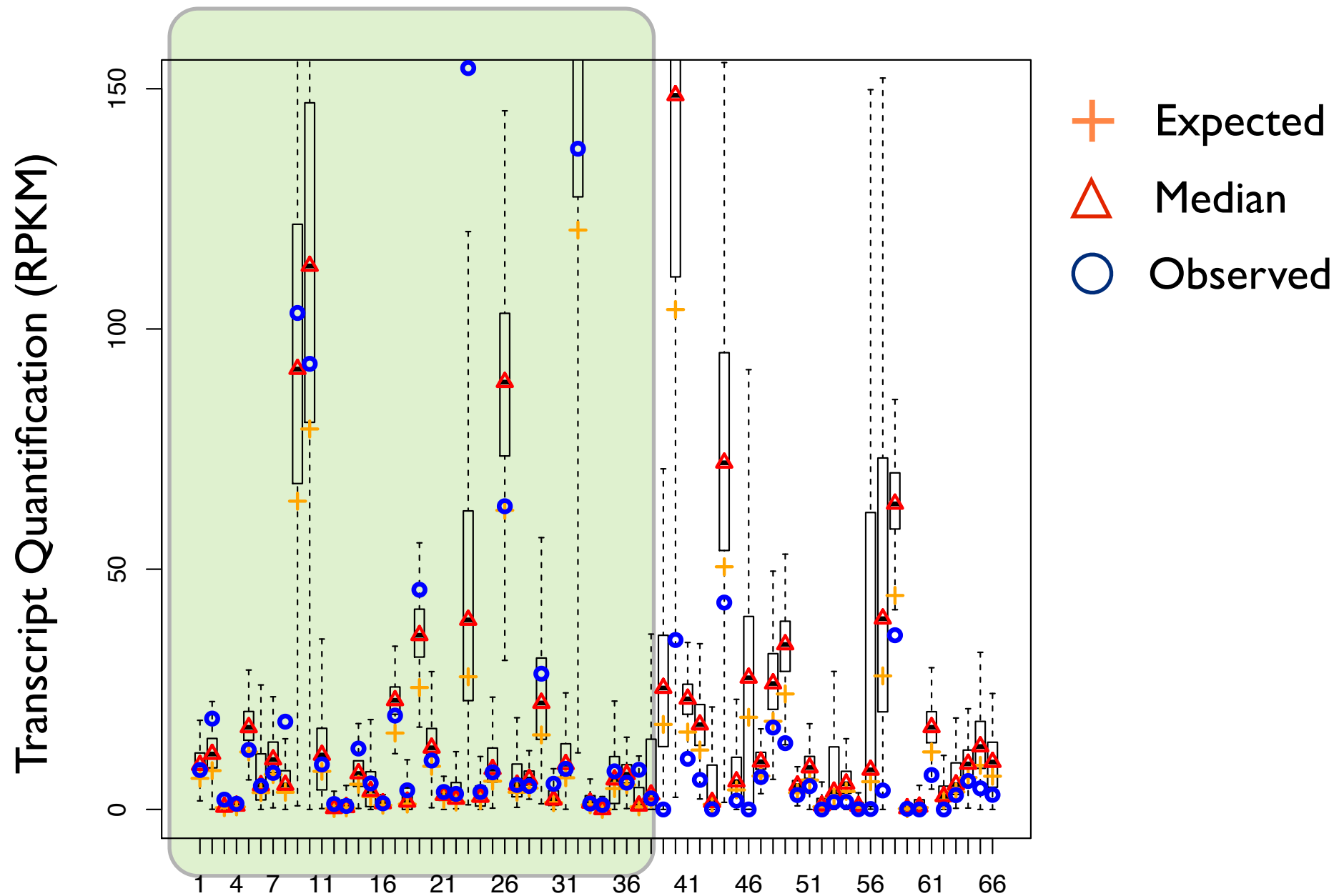
Dosage Compensation



Dosage Compensation

Dosage Compensation

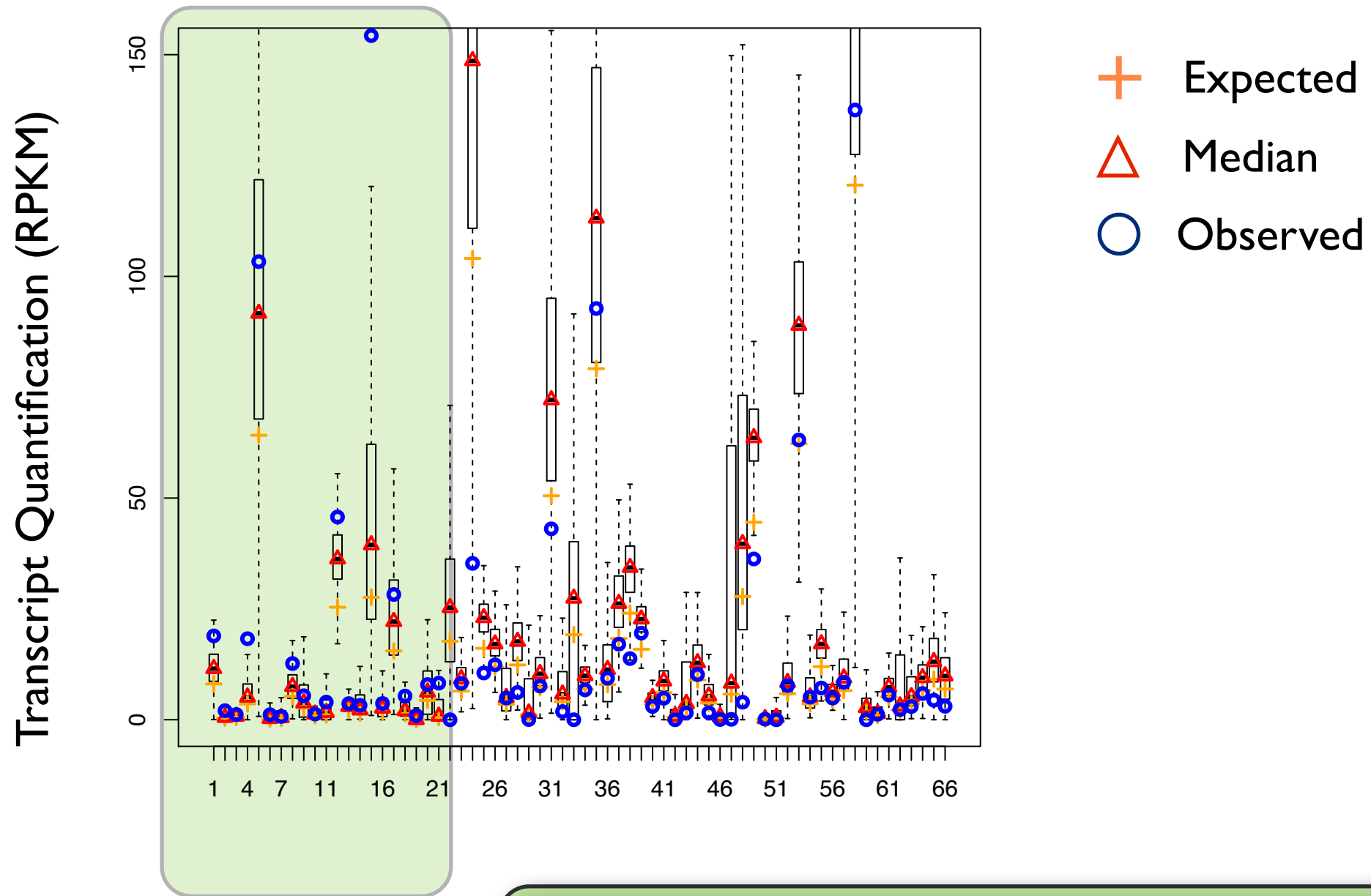
Dosage Larger
than expected



Dosage Compensation

Dosage Compensation

Dosage Larger
than median



Further investigation into Dosage
compensation (possible alternative
splicing events?)

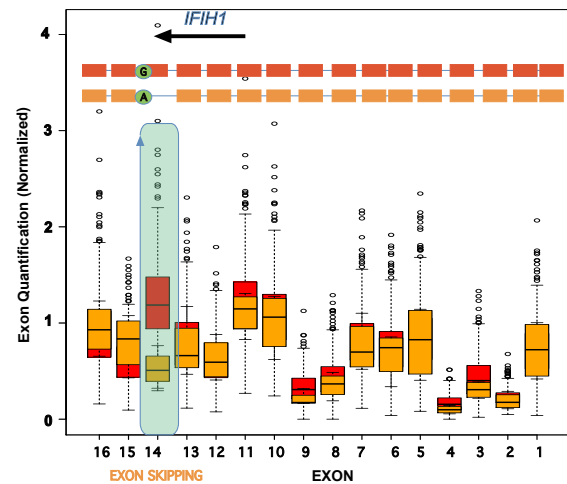
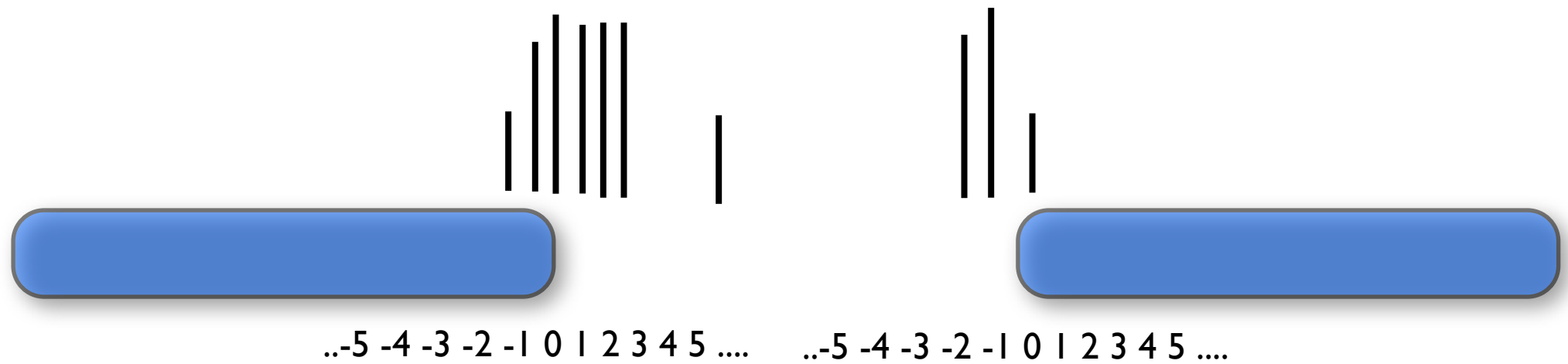
Dosage Compensation

Splice disrupting mutations

Exon Skipping

Intron Retention

Evaluate enrichment of positional motif signal associated to exon quantification

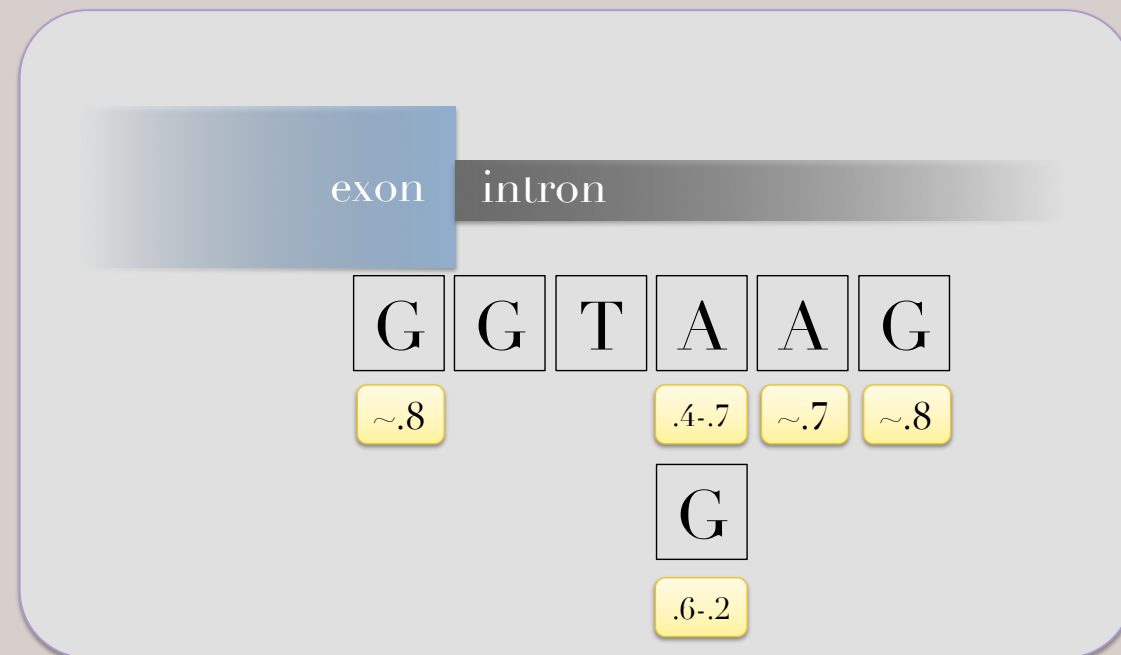


Variants impact Splicing

Rare Splice disrupting mutations

Creation of exonic 5'-splice-sites

5' splice sites

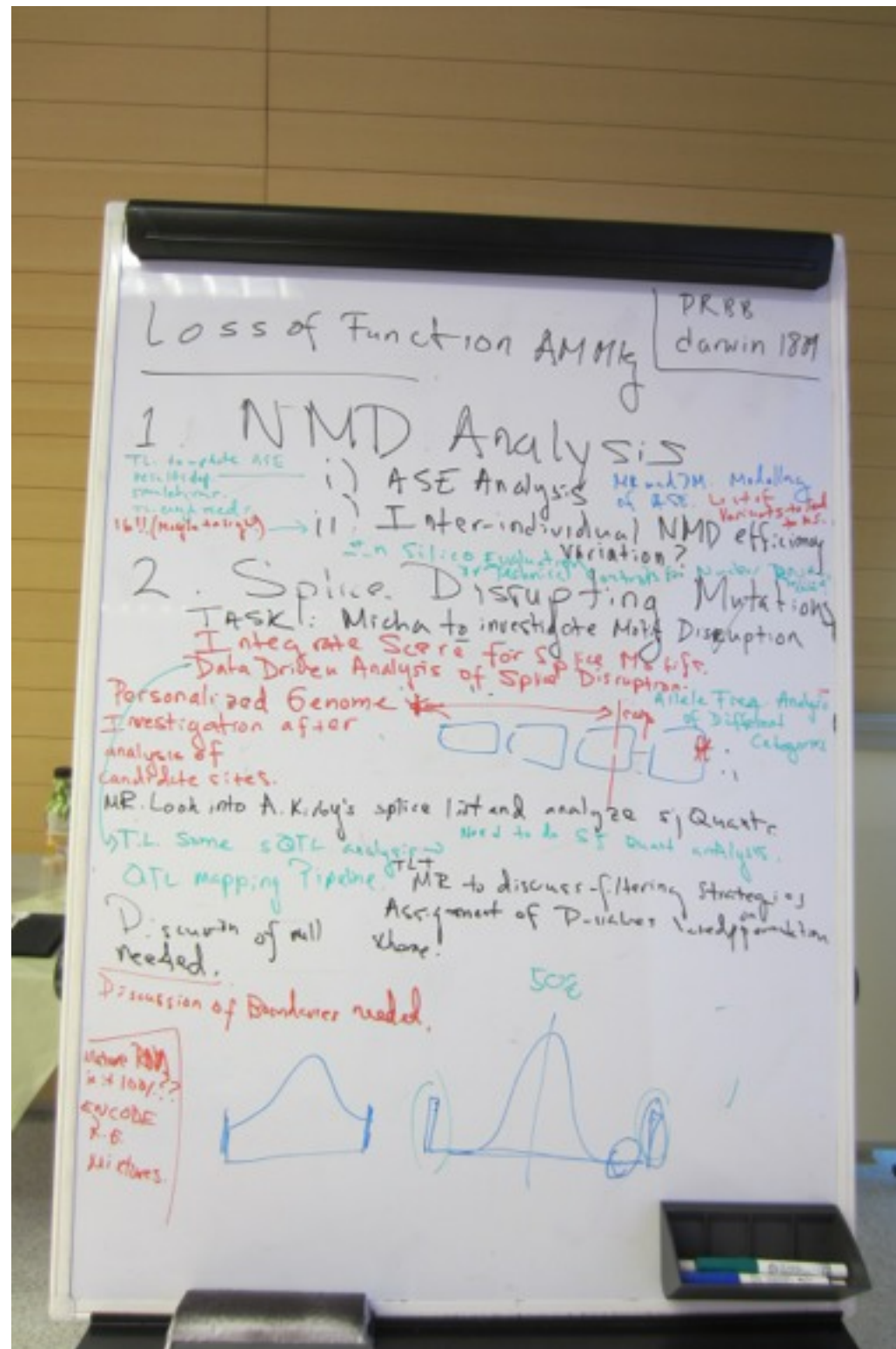


Hum Mol Genet. 1998 May;7(5):919-32. Statistical features of human exons and their flanking regions. Zhang MQ.

Some evidence based on allele frequency distribution in large exome sample sizes that these mutations are *deleterious*

Variants impact Splicing

TO DO:



lofgevadis@googlegroups.com